**AMS Common Exam Part B, Computational Biology Track, January Exam 2017**

**Name: _____ ID Num: _____**

**Part B: _____ / 75**

**DO THREE OUT OF FOUR QUESTIONS ONLY: Questions are based on AMS-535 (Questions 1-2) and CSE-549 (Questions 3-4). Each question is worth 25 points.**

## Question 1. Note this question has multiple parts.

**(1a)** Fill in the table with the three-letter codes and the corresponding one-letter codes for all naturally occurring amino acids and answer the following questions.

| No. | 3 letter code | one letter code | | No. | 3 letter code | one letter code |
|---|---|---|---|---|---|---|
| 1 | | | | 11 | | |
| 2 | | | | 12 | | |
| 3 | | | | 13 | | |
| 4 | | | | 14 | | |
| 5 | | | | 15 | | |
| 6 | | | | 16 | | |
| 7 | | | | 17 | | |
| 8 | | | | 18 | | |
| 9 | | | | 19 | | |
| 10 | | | | 20 | | |

**(1b)** Each amino acid has three components. What are they? Which component determines the characteristics of each amino acid?

_____

_____

**(1c)** Which side chain is smallest (three letter code)?

_____

**(1d)** Which side chains are charged (three letter codes)?

Positive charged = _____
Negative charged = _____

**(1e)** Which side chains are aromatic (three letter codes)?

_____

**(1f)** Indicate 4 non-charged residues capable of forming hydrogen bonds.

_____

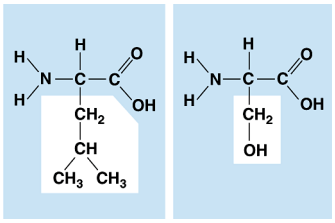
**(1g)** Which amino acid component projects outward from the axis of an alpha helix?

_____

**(1h)** Which side chain can form disulfide bonds (three letter code)?

_____

**(1i)** List by name the two amino acids shown below (three letter code) and indicate if they are hydrophilic or hydrophobic.



Left = _____

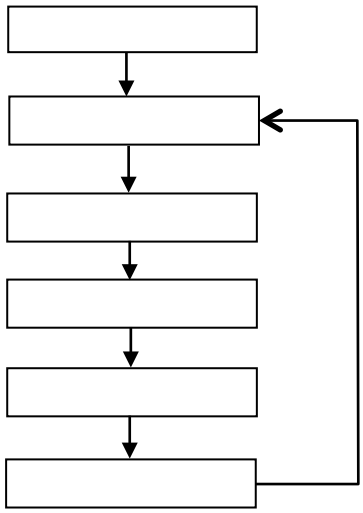
Right = _____

**(1j)** Write the three-letter code of the amino acid residue that has the ring incorporated into the peptide backbone which severely restricts the range of backbone torsion angles.

_____

**(1k)** Sketch the Lennard-Jones Potential in the following coordinate system and (i) clearly label the axes and include units, (ii) indicate the repulsive and attractive regions, and (iii) indicate the equilibrium point.

**(1l)** A typical genetic algorithm cycle includes six steps:, Breeding, New generation, Initial Generation, Survivors, Termination Check, and Fitness pressure (not necessarily in order). Fill in the following flow chart with these 6 terms in the correct order.



(**1m)** Write the most common functional form (i.e. the specific equations) for the classical potential energy functions used in computer simulations that employ a Molecular Mechanics force field. Explicitly label all variables and constants.

## Question 2. Note this question has multiple parts.

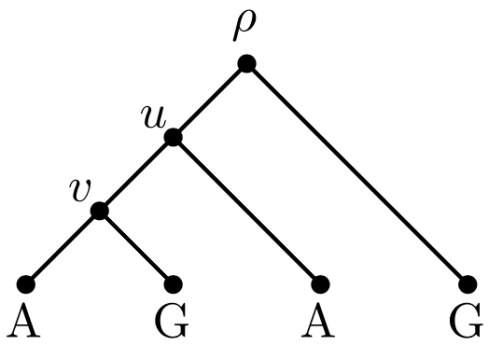
**(2a)** Explain in detail how Molecular Modeling force field equations and force field parameters were developed. Be specific and give detailed examples. Include in your answer particulars as to how the force fields were validated using experimental observables. What are the challenges for the future?

**(2b)** If a colleague asked you to collaborate on identification of drug-lead candidates for a newly discovered therapeutic protein target explain how you would go about performing a large-scale virtual screening experiment to prioritize compounds for purchase and experimental testing. Be specific and detailed in how you would go about the setups and the calculations. Include in your answer some potential problems (and remedies) that you might encounter.

**(2c**) Describe the differences between virtual screening and *de novo design*, give pros and cons of each method, and list at least three challenges associated with de novo growth of small organic molecules. Be detailed. How can challenges be overcome?

**Question 3: Note this question has multiple parts.**

**(3a)** Recall Fitch's algorithm for inferring parsimonious states in a phylogenetic tree. Consider the following tree,[1] and provide two solutions for the states of the internal nodes that will obtain the lowest cost (i.e., the maximum parsimony).




[1]Example from https://ab.inf.uni-tuebingen.de/teaching/ws2012/phylo/hw4.pdf

**(3b)** Devise a general algorithm that, under simple state transition costs (i.e., c(X, X) = 0, c(X, Y) = 1 if X != Y), will compute and return the total number of optimal solutions for the internal ancestral character states. Note, Fitch's algorithm *will not* find all possible solutions for the above tree, so you must modify the core algorithm in some way. Also note, your algorithm need not compute the optimal solutions themselves — only the number of such solutions.

**Question 4:**

You are given two strings S and T, and the suffix arrays of both — SA(S) and SA(T), respectively. Let the length of S and the length of T both be n. Assuming that the length of the largest substring shared by S and T is less than some number k, devise an O(nk) algorithm to construct the suffix array of S$T — that is, the generalized suffix array of S followed by T. You may either write your algorithm in pseudo-code, or describe it in clear, precise, and unambiguous language.